

АЛГОРИТМ ВИЛУЧЕННЯ ВЕКТОРІВ ОЗНАК В СИСТЕМІ ГОЛОСОВОЇ ІДЕНТИФІКАЦІЇ ОСОБИ

*Клеотюк В. О., магістрант; Дюжаєв Л. П., к.т.н., доц.
Національний технічний університет України «Київський політехнічний
інститут імені Ігора Сікорського», м. Київ, Україна*

В сучасному світі все більшого розвитку набувають біометричні системи (БС) контролю доступу користувача до певних ресурсів чи об'єктів. Головним показником БС є індивідуальна характеристика користувача, за якою надається право доступу. Одним із напрямів біометричної ідентифікації є ідентифікація людини за голосом.

Розглянемо алгоритм вилучення векторів ознак з вхідного мовного сигналу рис. 1.

Перший крок — перетворення аналогового сигналу в цифровий. Цей процес розбивається на два етапи: дискретизація і квантування. При дискретизації сигнал розбивається на окремі вимірювання амплітуди через деякий часовий інтервал. Число вимірювань в секунду називається частотою дискретизації. Процес подання дійсних значень цілими називається квантуванням. Надалі відцифрований сигнал буде позначатися як $x[t]$, де t деякий момент часу.

Після чого відбувається розбиття відцифрованого сигналу на невеликі фрагменти (фрейми), що характеризують окремі елементи мовного сигналу, для яких можна зробити припущення, що сигнал є стаціонарним. Розбиття на фрейми проводиться таким чином: вибирається віконна функція, яка приймає нульове значення всередині деякого часового відрізка, а за його межами — нульове. Далі вона послідовно застосовується до сигналу i , таким чином, «витягується» інформація всередині фрейму. Вилучення сигналу проводиться шляхом перемноження значення сигналу $x[t]$ в момент часу t зі значенням віконної функції $w[t]$ в момент часу t :

$$y[t] = w[t] \cdot x[t],$$

де $x[t]$ — значення сигналу; $w[t]$ — значення віконної функції.

Віконна функція характеризується трьома величинами: шириною (в мілісекундах), зміщенням (число мілісекунд між границями послідовних вікон) і формою.

Наступний крок — витяг спектральної інформації для сигналу, отриманого на попередніх етапах. На цьому етапі необхідно з'ясувати, яка кількість енергії міститься в кожному частотному діапазоні. Для вилучення спектральної інформації застосовується дискретне перетворення Фур'є.

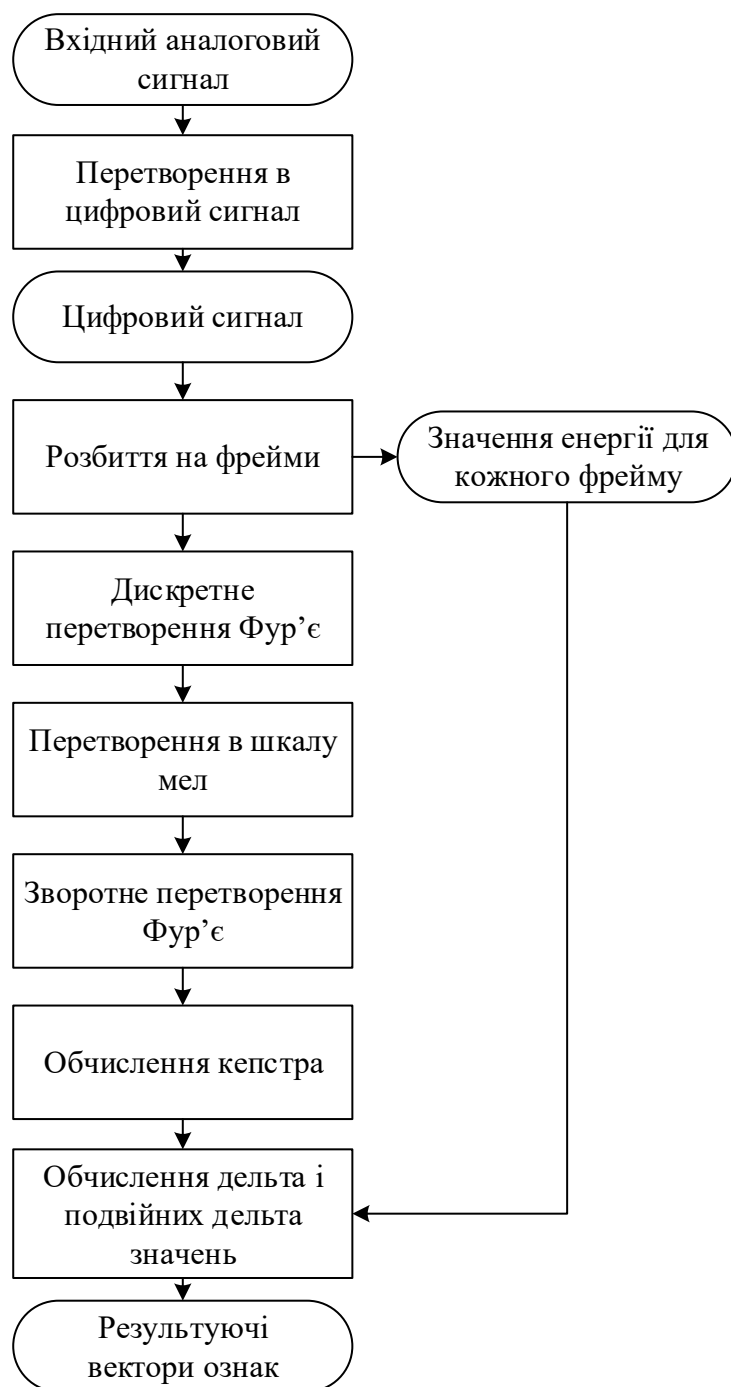


Рисунок 1. Алгоритм вилучення ознак з вхідного мовного сигналу

На вхід подається розбитий на фрейми сигнал, а на виході для кожного з T частотних діапазонів - комплексне число $X[k]$, яке є амплітудою і фазою вихідного сигналу:

$$X[k] = \sum_{t=0}^{T-1} x[t] e^{-i \frac{2\pi}{T} kt},$$

де $x[t]$ — значення сигналу; T — кількість частотних діапазонів.

Потім здійснюється перехід від величини частоти звуку до значення висоти (мел) за формулою:

$$mel(f) = 1127 \ln\left(1 + \frac{f}{700}\right),$$

де f — частота.

Це необхідно для врахування значущого чинника: людський слух має неоднакову чутливість для різних частотних діапазонів. Далі формуються трикутні фільтри, в яких накопичується 20 значень енергії для кожного з частотних діапазонів (10 фільтрів розподілені лінійно нижче 1000 Гц, а решта - логарифмічно вище 1000 Гц) і береться логарифм

кожного отриманого значення мел. Використання логарифма робить оцінки ознак менш чутливими до відмінностей у способі подачі вхідного сигналу.

Наступним кроком є обчислення кепстра. Кепстр дозволяє відокремити джерело звукової хвилі від фільтра, властивості якого проявляються при проходженні вихідної хвилі по голосовому каналу при утворенні звуку [1]. При цьому велика частина корисної інформації міститься в фільтрі.

Формально кепстр можна визначити наступним чином:

$$c[t] = \sum_{t=0}^{T-1} \log \left(\left| \sum_{t=0}^{T-1} x[t] e^{-i \frac{2\pi}{T} kt} \right| \right) e^{-i \frac{2\pi}{T} kt},$$

де $x[t]$ — значення сигналу; T — кількість частотних діапазонів;

Крім того, для кожного фрейму додається ще одна ознака — енергія сигналу:

$$E = \sum_{t=1}^{t2} x^2[t].$$

Останнім етапом є обчислення дельта значень і подвійних дельта значень для отриманих на попередніх етапах ознак, які також додаються в результуючий вектор ознак. Дельта значення обчислюються за формулою:

$$d(t) = \frac{c(t+1) - c(t-1)}{2},$$

де $c(t)$ — кепстральний коефіцієнт.

Подвійні дельта значення обчислюються аналогічно, з тією лише різницею, що замість кепстральних коефіцієнтів використовуються обчислені дельта значення [2].

Такий алгоритм вилучення векторів ознак є простим в реалізації, але потребує прийнятних обчислювальних потужностей при використанні його у системах голосової ідентифікації.

Перелік посилань

1. Дюжаєв Л. П. Аналіз впливу параметрів обробки звукового сигналу на якість розпізнавання голосових команд / Л. П. Дюжаєв, В. Ю. Коваль // Вісник НУТУ «КПІ». Сер. Радіотехніка. Радіоапаратобудування. — 2014. — №56. — С. 34—41.

2. Розпізнавання голосу — Режим доступу: https://ru.wikipedia.org/wiki/Распознавание_голоса — Назва з екрану

Анотація

В роботі представлено ефективний алгоритм для вилучення векторів ознак з вхідного мовного сигналу в системі голосової ідентифікації особи.

Ключові слова: голосова біометрія, біометрична система, розпізнавання голосу.

Аннотация

В работе представлен эффективный алгоритм, касающийся извлечения векторов признаков из входного речевого сигнала в системе голосовой идентификации личности.

Ключевые слова: голосовая биометрия, биометрическая система, распознавания голоса.

Abstract

This article presents the effective algorithm dealing with the extracting features vectors of the input speech signal in voice identification.

Keywords: voice biometrics, biometric system, voice recognition.